ESTIMATING WHAT IS UNDER THE TIP OF THE GENDER-BASED VIOLENCE ICEBERG: A STATISTICAL MODELLING APPROACH

D. Moriña, I. Millán, A. Fernández-Fontelo, P. Toran, P. Puig, V. Alonso-Lara, M. Cobo, M. Arias, M. Bolívar, N. Pascual and G. Falguera

Gender-based violence

- Gender Based Violence (GBV) is one of the current primary public health concerns world-wide. According to the World Health Organization (WHO), 30% of women worldwide have been subjected to either physical and/or sexual intimate partner violence or non-partner sexual violence in their lifetime
- In Spain, 21.5% of women aged 16 or older have suffered physical violence throughout their lives and 13.7% have suffered sexual violence throughout their lives

Gender-based violence

- However, this report also found that only 32% of the victims of physical and sexual GBV report the violent episodes to the police or to the court
- In order to deal adequately with the GBV problem we should be able to estimate and reach all the cases of GBV under the "iceberg". One of the first questions that arise within this issue is why the female victims of GBV do not report the violent events

4 Gender-based violence

- The reason so many cases go unreported are both personal (embarrassment, fear of retaliation, economic dependency) and societal (imbalanced power relations for men and women in society, privacy of the family, victim blaming attitudes)
- Many of these reasons carry a social stigma and have been cultural heritage for hundreds or thousands of years. Another question that arises is if the submerged part of the iceberg is a matter of ignorance or a matter of social silence and inhibition

5 Training activities

- The health-sector actions in Catalonia are concentrated on creating protocols at the various levels of health and social care and inter-sector coordination
- In 2019, an intervention is conducted to raise awareness among the health professionals

COVID-19 pandemic

- In March 2020, the state of emergency declared by WHO due to COVID-19 had as consequence that Spain among other countries took a drastic measure of confinement of the whole population from March 14 to June 21

- The home confinement caused by SARS-COV-2 pandemic was a risk factor for GBV as the economic, employment and stress problems increased and there was a prolonged coexistence between victims and aggressors

- A recent analysis showed that the number of calls to Spain's helpline for GBV (dial 016) had an important and evident increase since March up till June 2020 compared with the previous two years

Objectives

- Evaluate the impact of the 2019 campaign
- Evaluate the impact of the COVID-19 pandemic in the number of GBV events
- Quantify the magnitude of unreported GBV cases in the context of the Catalan public primary care system
- Estimate the real evolution of GBV events specifically in the northern metropolitan area of Barcelona in Catalonia from 2010 to 2021

Data

Three data sources are available:

- Weekly number of GBV-related diagnoses in each administrative area of the northern metropolitan area of Barcelona in Catalonia from 2010 to 2021
- Number of women assigned to each administrative area
- Macro survey on GBV conducted by the Spanish Ministry of Equity in 2019

Data

From the 2019 spanish survey we can learn that 32% of women in Barcelona province have suffered physical and / or sexual violence episodes but only 16% seek for any health attention, and therefore we expect:

- Subarea A: 84,232 women. $\hat{\lambda}_A = 7$
- Subarea B: 91,639 women. $\hat{\lambda}_B = 7$
- Subarea C: 52,968 women. $\hat{\lambda}_C = 4$
- Subarea D: 168,783 women. $\hat{\lambda}_D = 13$
- Subarea E: 4,695 women. $\hat{\lambda}_E = 1$
- Subarea F: 184,594 women. $\hat{\lambda}_F = 14$

The model

Let's assume that the actual weekly number of GBV cases X_t follows a Poisson distribution with mean λ, which is increased in a factor β in the mandatory confinement period (2020 March 14th to 2020 June 24th), i.e., $E(X_t) = \lambda + I(t) \cdot \beta$ where $I(t)$ takes the value 1 if t falls within the mandatory confinement period and 0 otherwise

The model

The number of cases diagnosed within the public primary care system, Y_t, is just a part of the actual process, expressed as

$$Y_t = \begin{cases} q_0 \circ X_t, t \le t' \\ q_t \circ X_t, t > t' \end{cases} \tag{1}$$

where $\circ$ is the *binomial thinning* operator, defined as $q_t \circ X_t = \sum_{i=1}^{X_t} Z_i$, with Z_i independent and identically distributed Bernoulli random variables with probability of success q_t and $q_t = q_0 + \frac{t-t'}{\alpha-t'} \cdot (1 - q_0)$ for $t > t'$

The model

Additionally, the impact of the training on the underreporting process has been modelled with an exponential shape instead of the linear:

$$q_t = 1 - (1 - q_0) \cdot e^{-\alpha \cdot (t - t')} \tag{2}$$

For each administrative area we choose the best fitting approach (according to Deviance Information Criterion)

13   The model

It should also be noted that α is the moment when $q_\alpha = 1$, i.e., the registered and observed processes coincide. All the parameters (q_0, λ, β, α and t') are estimated by Gibbs sampling using the *R2jags* package, using appropriate priors based on the available information

Priors

In order to assure identifibility of the process, we need to set additional assumptions on the prior for λ:

- Scenario 1: λ prior is based on 2019 national macrosurvey estimates (32% of women are victims of GBV at some point in their lifes and only 16% seeks primary health attention afterwards)

- Scenario 2: λ prior is based on the assumption that 50% of GBV victims seek primary health attention afterwards

- Scenario 3: λ prior is based on the assumption that 90% of GBV victims seek primary health attention afterwards

Best modelling approach for q_t

Subarea	Linear	Exponential
A	1987.3	1979.7
B	2293.6	2289.5
C	1361.9	1359.0
D	3266.5	3251.5
E	2051.1	1830.2
F	2596.7	2572.4

Impact of the training (linear assumption)
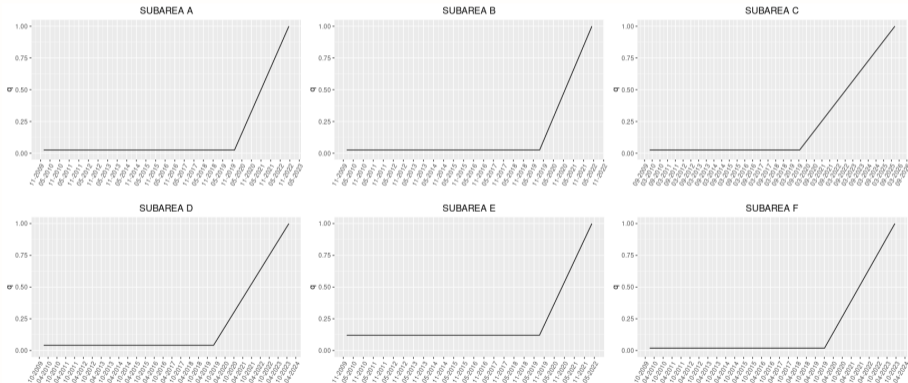
17   Impact of the training (exponential assumption)
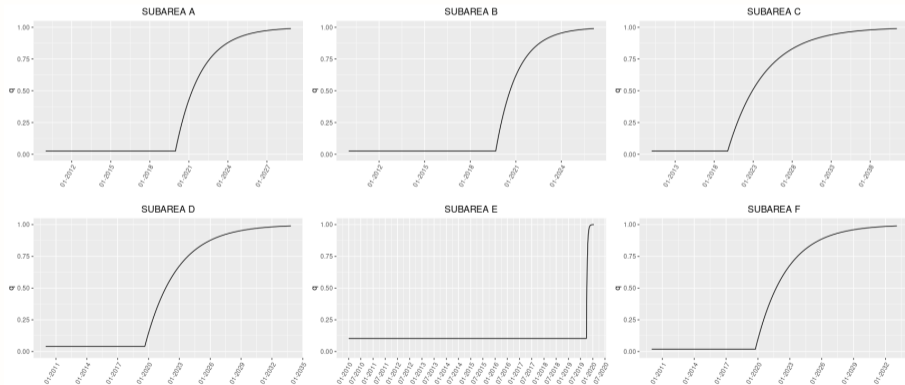
18   Impact of the training (exponential assumption)

	Dâte
A	2028-11-03
B	2026-02-20
C	2041-05-31
D	2033-11-04
E	2019-11-22
F	2033-01-28

19   Impact of the training (exponential assumption)
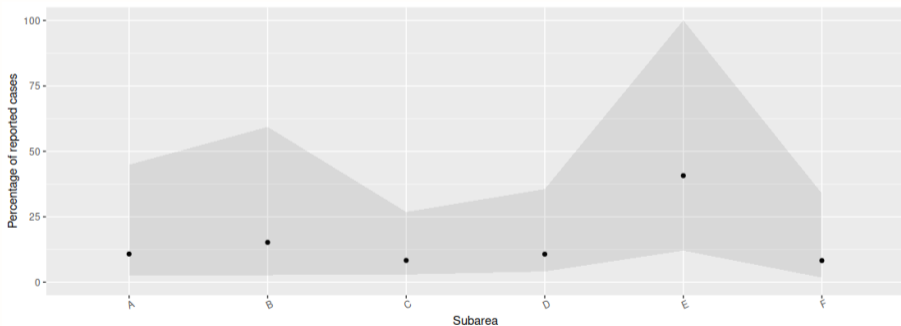
20   Impact of the training (exponential assumption) - Scenario 2
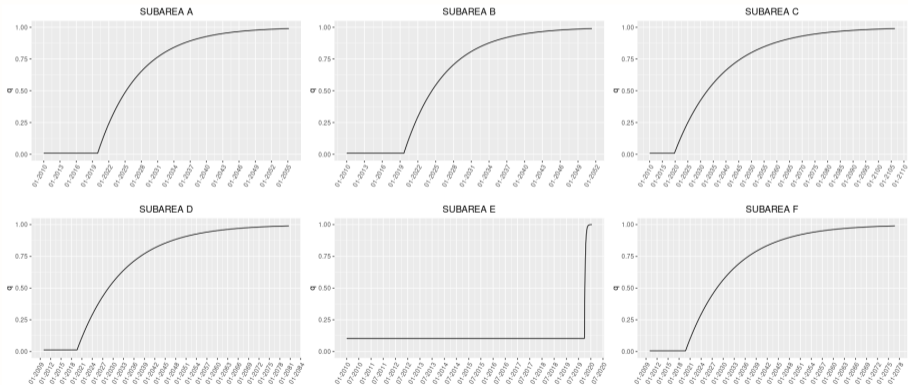
21   Impact of the training (exponential assumption) - Scenario 2

	Dâte
A	2055-03-19
B	2051-05-12
C	2106-06-11
D	2080-08-09
E	2019-11-22
F	2076-09-25
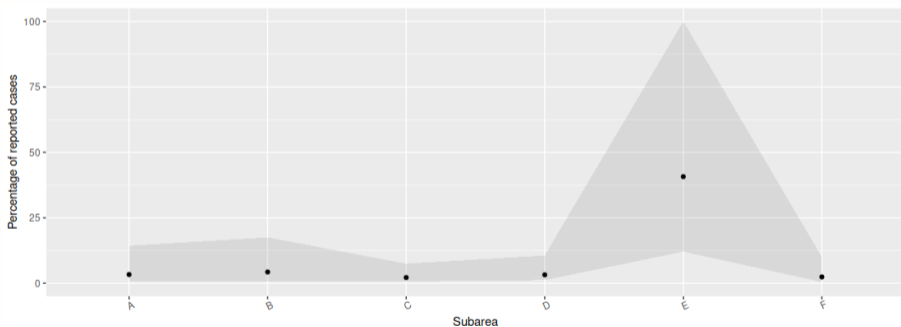
Impact of the training (exponential assumption) - Scenario 2
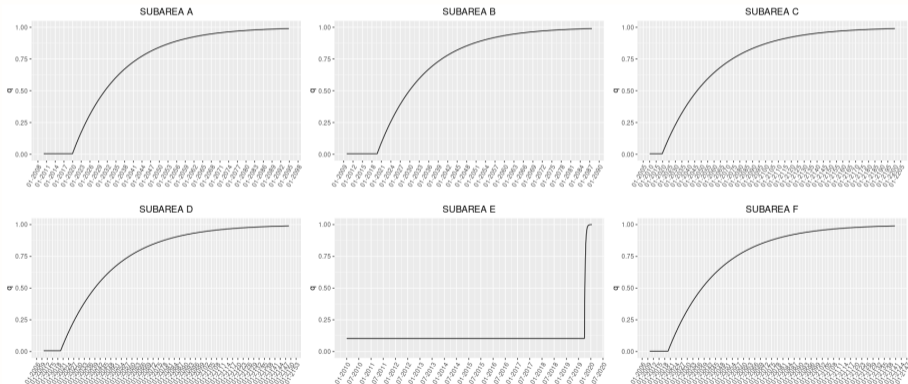
Impact of the training (exponential assumption) - Scenario 3

	Dâte
A	2094-11-05
B	2087-09-19
C	2200-06-20
D	2148-11-29
E	2019-11-22
F	2140-08-19

25   Impact of the training (exponential assumption) - Scenario 3
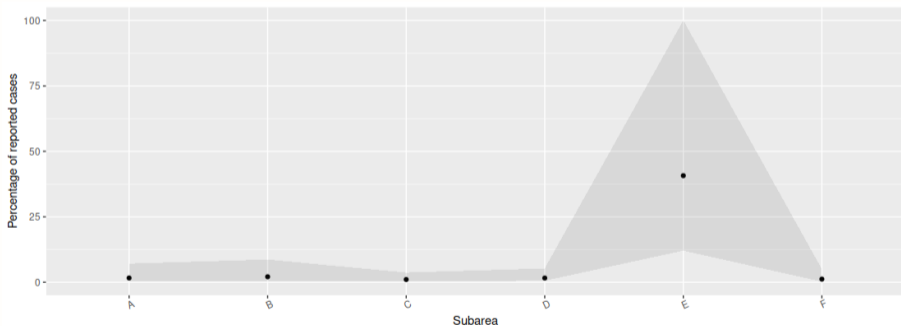
26   Underdiagnosing before the training (exponential assumption)

	$\hat{q_0}$	$\hat{\lambda}$
A	0.024 (0.020, 0.030)	7.0 (6.8, 7.2)
B	0.026 (0.021, 0.031)	7.0 (6.8, 7.2)
C	0.025 (0.019, 0.033)	4.0 (3.8, 4.2)
D	0.041 (0.036, 0.047)	13.0 (12.8, 13.2)
E	0.103 (0.079, 0.134)	1.4 (1.2, 1.5)
F	0.018 (0.015, 0.022)	14.0 (13.8, 14.1)

Underdiagnosing before the training (exponential assumption) - Scenario 2

	$\hat{q_0}$	$\hat{\lambda}$
A	0.008 (0.007, 0.01)	21.0 (20.8, 21.2)
B	0.008 (0.006, 0.01)	23.0 (22.8, 23.2)
C	0.008 (0.006, 0.01)	13.0 (12.8, 13.2)
D	0.013 (0.011, 0.015)	42.0 (41.8, 42.1)
E	0.103 (0.079, 0.134)	1.4 (1.2, 1.5)
F	0.006 (0.005, 0.007)	46.0 (45.8, 46.2)

Underdiagnosing before the training (exponential assumption) - Scenario 3

	$\hat{q_0}$	$\hat{\lambda}$
A	0.004 (0.003, 0.005)	42.0 (41.8, 42.2)
B	0.004 (0.003, 0.005)	46.0 (45.8, 46.2)
C	0.004 (0.003, 0.005)	26.0 (25.8, 26.2)
D	0.006 (0.006, 0.007)	84.0 (83.8, 84.2)
E	0.103 (0.079, 0.134)	1.4 (1.2, 1.5)
F	0.003 (0.002, 0.003)	92.0 (91.8, 92.2)

Impact of the Covid-19 confinement

	$\hat{\beta}$
A	1.3 (0.2, 3.7)
B	0.8 (0.1, 2.4)
C	1.4 (0.2, 3.9)
D	1.8 (0.3, 2.7)
E	0.9 (0.2, 2.0)
F	1.3 (0.2, 3.8)

Impact of the Covid-19 confinement - Scenario 2

	$\hat{\beta}$
A	2.0 (0.3, 6.0)
B	1.5 (0.2, 4.7)
C	1.9 (0.3, 5.7)
D	2.1 (0.3, 6.6)
E	0.9 (0.2, 1.3)
F	1.8 (0.3, 5.7)

Impact of the Covid-19 confinement - Scenario 3

	$\hat{\beta}$
A	1.9 (0.3, 6.2)
B	1.6 (0.2, 5.3)
C	1.8 (0.3, 5.9)
D	1.9 (0.3, 6.3)
E	0.9 (0.2, 2.0)
F	1.8 (0.3, 5.8)
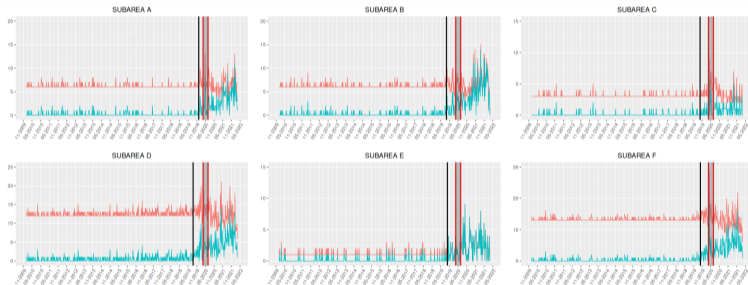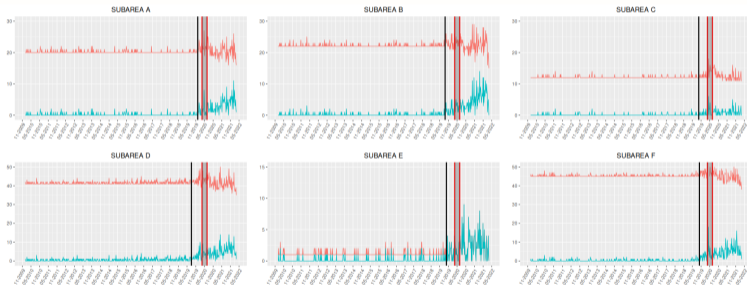
Reconstructed process (exponential assumption)

Reconstructed process (exponential assumption) - Scenario 2
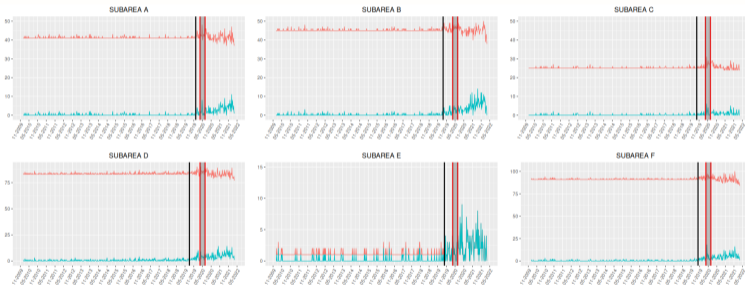
34   Reconstructed process (exponential assumption) - Scenario 3
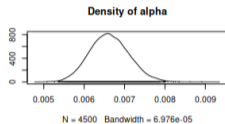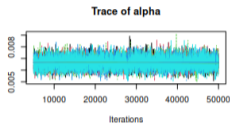
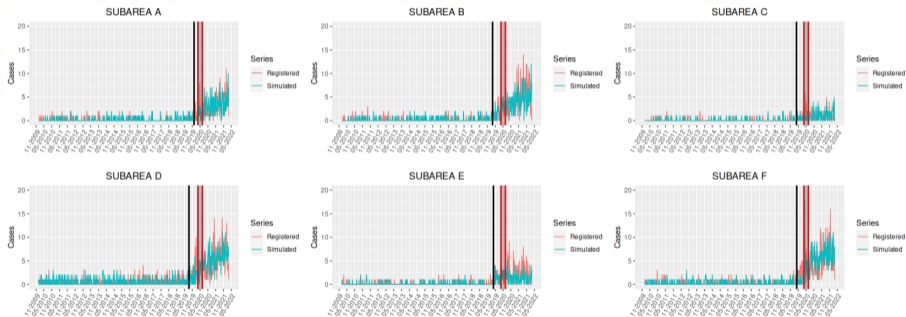# Model diagnosis (exponential assumption)
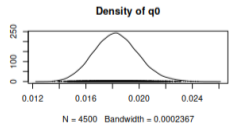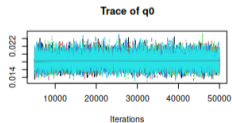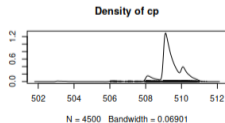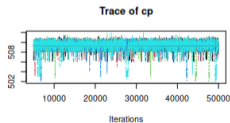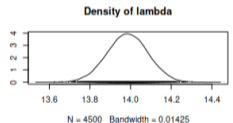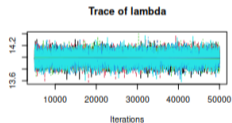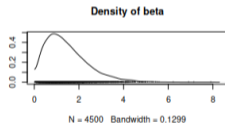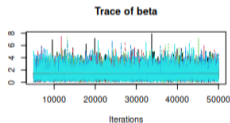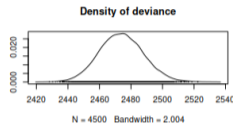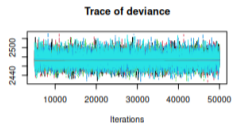
Simulated process (exponential assumption)

- We considered an exhaustive simulation study reproducing the described structure with different parameter values has been conducted in order to assess whether the original values can be recovered by using this estimation method and to evaluate the model performance, both for the linear and exponential modelling of q_t.

| q_t	Parameter	Bias (%)	AIL	Coverage (%)
Linear	α	1.91	2207	90.1
	β	0.002	3.98	96.4
	t'	0.3	495	98.2
	λ	0.0002	0.38	89.3
	q_0	0.00004	0.07	89.1
Exponential	α	0.0002	0.93	99.4
	β	0.002	4.19	97.4
	t'	0.089	155	98.0
	λ	0.0002	0.35	95.5
	q_0	< 0.0001	0.083	93.6

Conclusions

- Although some data science approaches have been recently proposed in the context of gender-based violence, most are mainly focused on identifying risk factors related to it
- The results of this work show that, in average, only around 16% of the gender-based violence cases occurred between 2010-01-01 and 2021-12-31 in the North Metropolitan Health area that requested primary care attention were properly detected and registered

Conclusions

- The severity of the underreporting issue was particularly dramatic at the beginning of the time period, before training the professionals in charge of diagnosing these cases, when only 4% of the cases was registered. After the training, an average of 50% of the cases was registered
- The impact of this sensibilization activity can therefore be quantified as this `11`-fold reduction in the underreporting issue

Conclusions

- Another important result that can be learnt from the present study is the quantification of the impact of the mandatory confinement due to the COVID-19 pandemic (March-June 2020), estimated as an average growth of 1.25 cases per week
- The results reported in this work are not exclusive of the specific geographic area for which actual data was analyzed

Further work

Further work
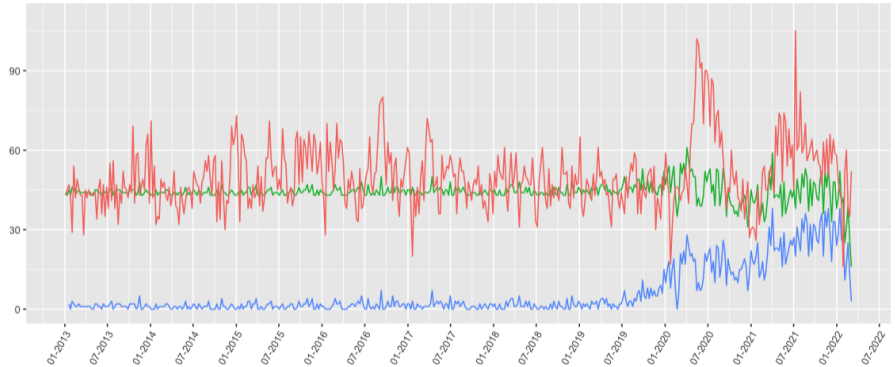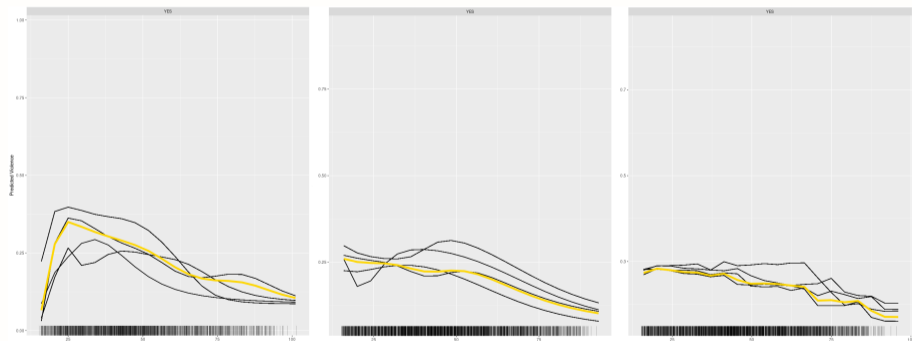
More information

More information in the project websites:
`https://dmorinya.github.io/publications/`
`https://dmorina.shinyapps.io/DaSciVioDesc/`

Gràcies!   dmorina@ub.edu